2-Day Course – Spatial Modeling with Geostatistics

Prof. Michael J. Pyrcz, Ph.D., P.Eng.
Associate Professor

Hildebrand Department of Petroleum & Geosystems Engineering
University of Texas at Austin

Bureau of Economic Geology, Jackson School of Geosciences
University of Texas at Austin

“In two days, what a geoscientists needs to know about geostatistics, and workflows to get you started with applying geostatistics to impact your work.”

Spatial Modeling with Geostatistics

Spatial Data Analysis - Modeling

Lecture outline . . .

- Estimation and Simulation
- Sequential Gaussian Simulation

Prerequisites

Introduction

Probability Theory

Representative Sampling

Spatial Data Analysis

Spatial Estimation

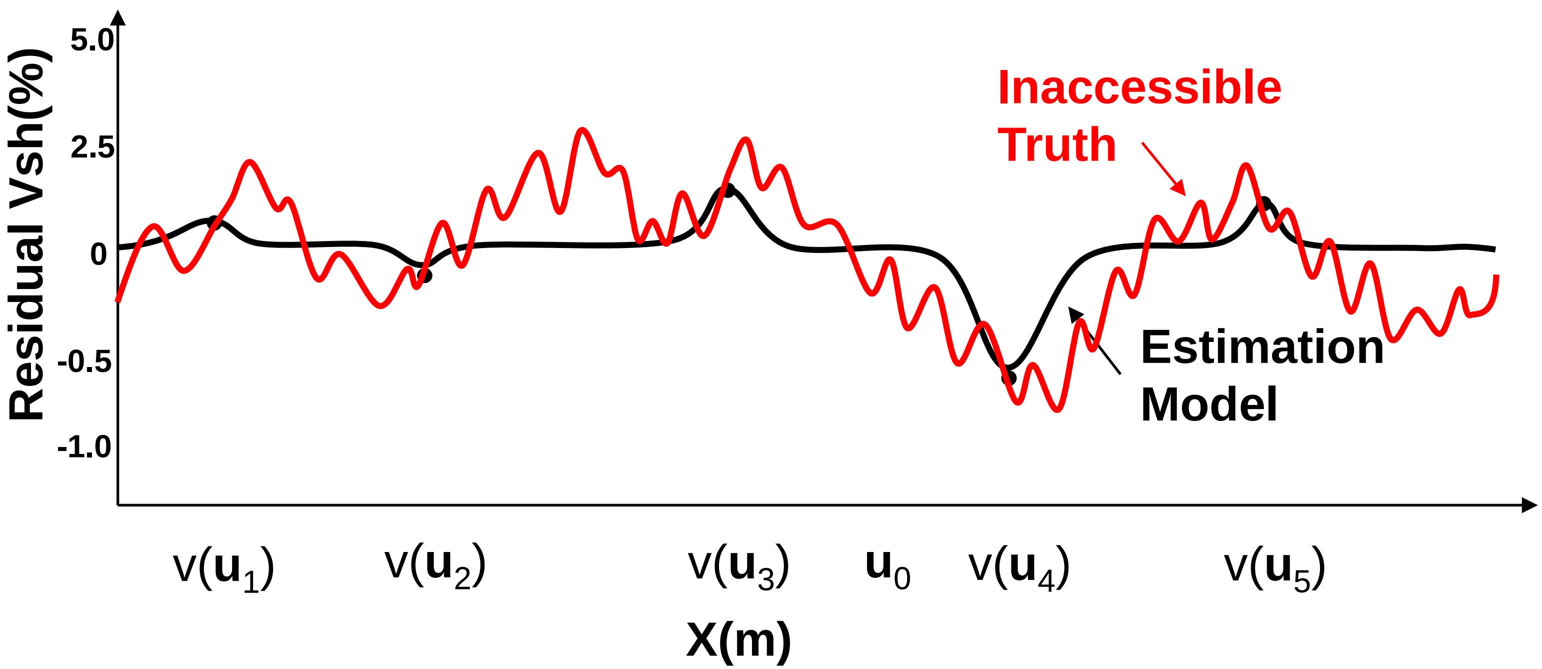
Stochastic Simulation

Uncertainty Management

Machine Learning

What's Wrong with Kriging?

- Kriging is an estimation method. The goal of estimation is the most likely value at each location. Too smooth!
  - Would kriging honor the global distribution? Variance is too low!
  - Would kriging honor the variogram model? Super continuous!



  - We need a method that corrects for the smoothing ⇨ simulation

Estimation vs. Simulation Definition

Estimation:

- honors local data (with discontinuity)
- locally accurate
- smooth appropriate for visualizing trends
- inappropriate for flow simulation
- no assessment of global uncertainty

Simulation:

- honors local data
- reproduces histogram
- honors spatial variability → appropriate for flow simulation
- alternative realizations possible → change random number seed
- assessment of global uncertainty is possible

Smoothing Effect of Kriging

- Kriging is locally accurate and smooth, appropriate for visualizing trends, inappropriate for process evaluation where extreme values are important

- The "variance" of the kriged estimates is too small:

$$Var\{Y^*(\mathbf{u})\} = \sigma^2 - \sigma^2_{SK}(\mathbf{u})$$

Simple Kriging Estimation Variance

  – σ^2 is complete variance

  – $\sigma^2_{SK}(\mathbf{u})$ is zero at the data locations → no smoothing

  – $\sigma^2_{SK}(\mathbf{u})$ is variance σ^2 far away from data locations → complete smoothing

  – spatial variations of $\sigma^2_{SK}(\mathbf{u})$ depend on the variogram and data spacing

Smoothing Effect of Kriging

- Missing variance *is* the kriging variance $\sigma_{SK}^{2}(\mathbf{u})$

- The idea of *simulation* is to correct the variance (get right histogram) and get the right variogram

$$Y_s(\mathbf{u}) = Y^*(\mathbf{u}) + R(\mathbf{u})$$

where *R*(u) corrects for the missing variance.

- Simulation reproduces histogram, honors spatial variability (variogram) → appropriate for process evaluation

- Allows an assessment of uncertainty with alternative realizations

Sequential Simulation

- Transform data to standard normal distribution (all work will be done in "normal" space)
- Go to a location and perform kriging to get mean and corresponding kriging variance:

$$Y^*(\mathbf{u}) = \sum_{\beta=1}^{n} \lambda_\beta \cdot Y(\mathbf{u}_\beta)$$

$$\sigma_{SK}^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n} \lambda_\alpha C(\mathbf{u},\mathbf{u}_\alpha)$$

- Draw a random residual *R*(u) that follows a normal distribution with mean of 0.0 and variance of $\sigma_{SK}^2(\mathbf{u})$
- Add the kriged estimate and residual to get simulated value:

$$Y_s(\mathbf{u}) = Y^*(\mathbf{u}) + R(\mathbf{u})$$

- Note that Y_s(u) could be equivalently obtained by drawing from a normal distribution with mean *Y**(u) and variance $\sigma_{SK}^2(\mathbf{u})$

Sequential Simulation

- Add Y_s(u) to the set of data to ensure that the covariance with this value and all future predictions is correct
- A key idea of sequential simulation is to use previously kriged/simulated values as data so that we reproduce the covariance between *all* of the simulated values!
- Visit all locations in random order (to avoid artifacts of limited search)
- Back-transform all data values and simulated values when model is populated
- Create another equiprobable realization by repeating with different random number seed

Why Gaussian Simulation?

- Local estimate is given by kriging
- Mean of residual is zero and variance is given by kriging; however, what "shape" of distribution should we consider?
- Advantage of normal / Gaussian distribution is that the global N(0,1) distribution will be preserved if we always use Gaussian distributions
- Could derive theory in terms of the multivariate Gaussian distribution
- Transform data to normal scores in the beginning (before variography)
- Simulate 3-D realization in "normal space"

Why Gaussian Simulation?

- Back-transform all of the values when finished
- Price of mathematical simplicity is the characteristic of maximum spatial entropy, i.e., low and high values are disconnected. Not appropriate for permeability.
- Consequences:
  - maximum spatial disorder beyond the variogram
  - maximum disconnectedness of extreme values
  - median values have greatest connectedness
  - symmetric disconnectedness of extreme low / high values

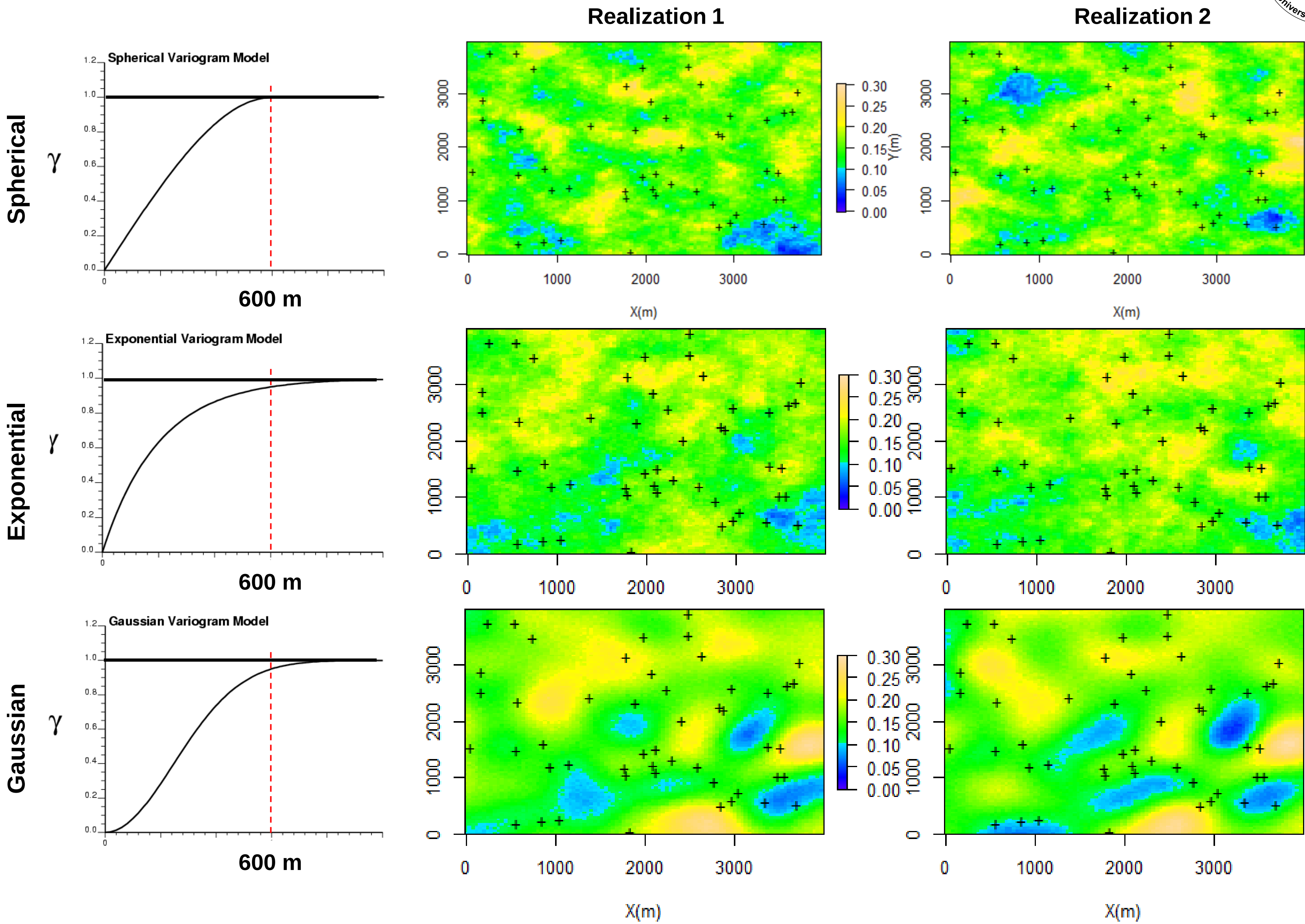
Steps in Sequential Gaussian Simulation

1. Transform data to “normal space”
2. Establish grid network and coordinate system (-space)
3. Assign data to the nearest grid node (take the closest of multiple data assigned to the same node)
4. Determine a random path through all of the grid nodes
   - » find nearby data and previously simulated grid nodes
   - » construct the conditional distribution by kriging
   - » draw simulated value from conditional distribution
5. Check results
   - » honor data?
   - » honor histogram: N(0,1) standard normal with a mean of zero and a variance of one?
   - » honor variogram?
   - » honor concept of geology?
6. Back transform

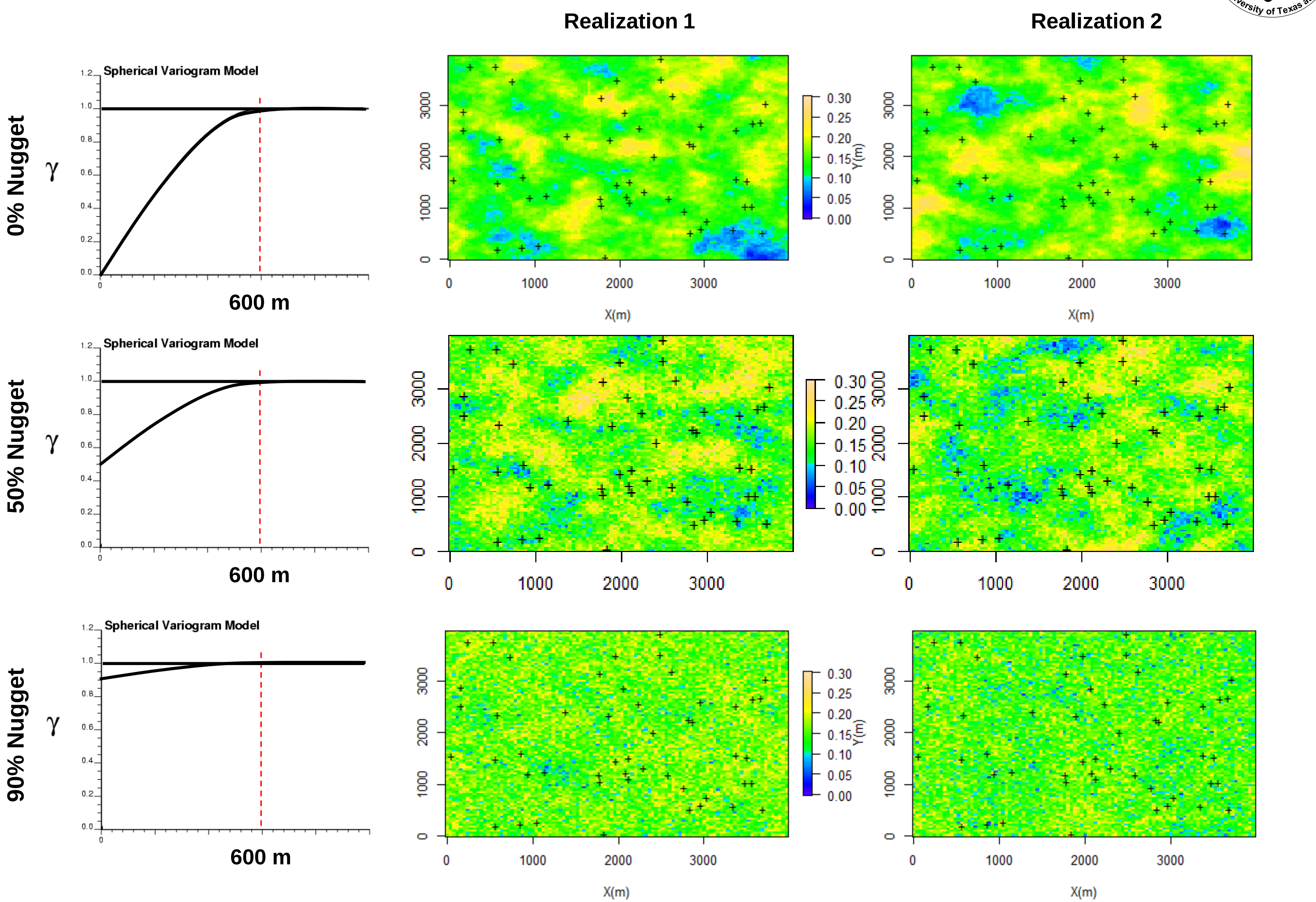
Sequential Gaussian Simulation Definition

- **Sequential** – sequential inclusion of simulated values to impose the correct spatial correlation between the simulated values.
- **Gaussian** – work in Gaussian space since the local conditional distribution shape is known and can be parameterized by mean (kriging estimate) and variance (estimation variance).
- **Simulation** – simulation through Monte Carlo simulation from the local distribution of uncertainty and construction of multiple, equiprobable realizations.

Some SGSIM Realizations

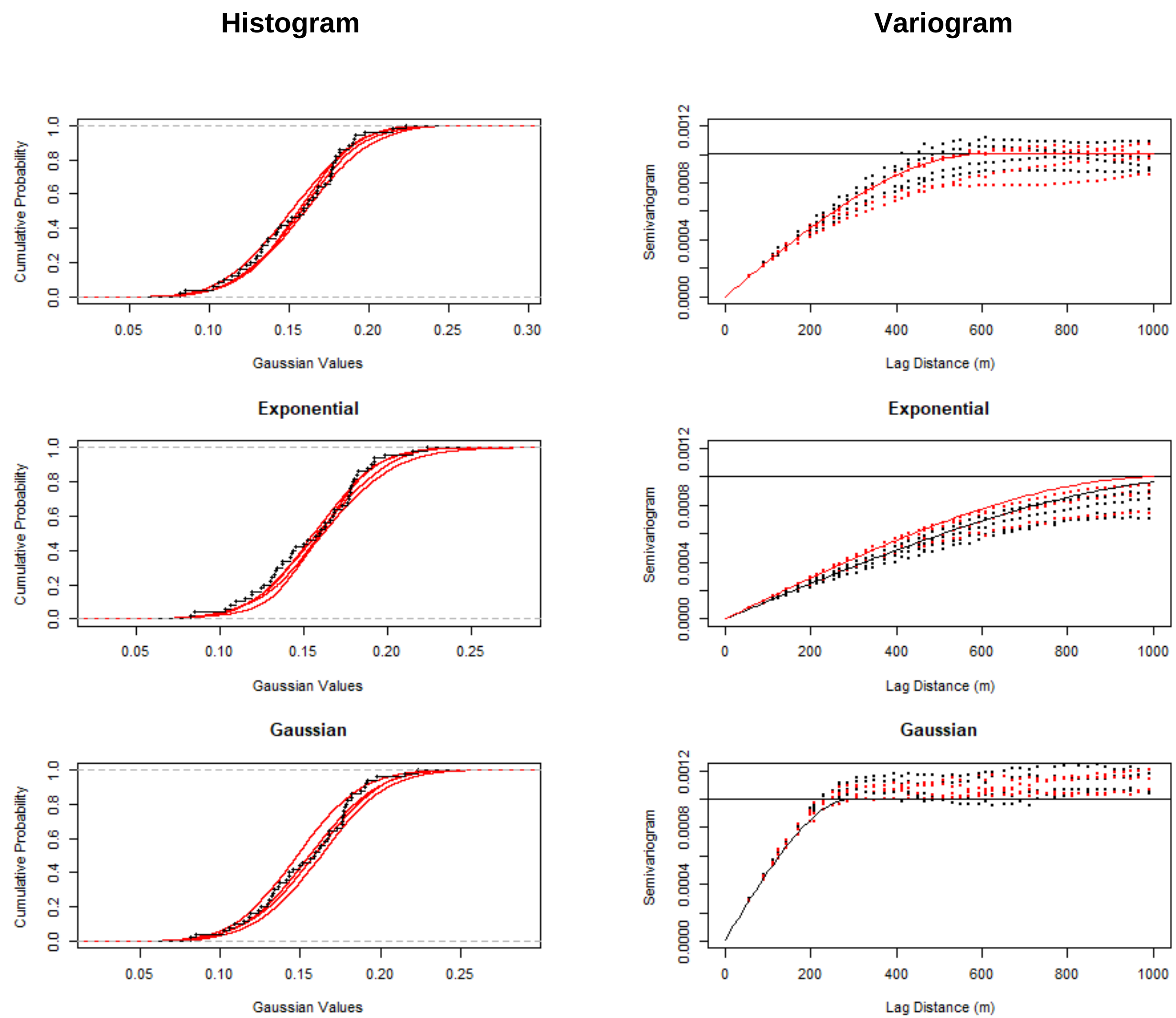


Some SGSIM Realizations



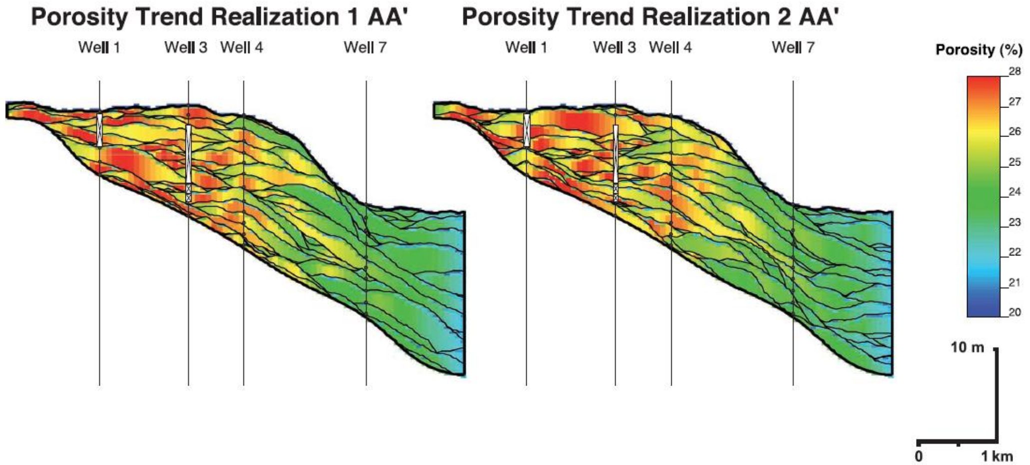
Ergodic Fluctuations

- Expect some statistical fluctuation in the input statistics
- These are a function of the ratio of spatial continuity to the size of the model.
  - If model is large relative to spatial continuity range then fluctuations should minimal
  - If model is small relative to spatial continuity range then fluctuations may be extreme



Other Types of Simulation

- Categorical Simulation for Facies
  - Truncated Gaussian Simulation
    - » Bin the continuous simulation method
  - Sequential Indicator Simulation
    - » Use of indicator variograms
  - Multiple-point Simulation
    - » With training Images instead of variograms
  - Object-based Simulation
    - » Stochastic placement of geometric
  - Rule-based, Process Mimicking
    - » Stochastic forward modeling



Rule-based deepwater lobes reservoir model (Pyrcz et al., 2005).

Gaussian Simulation Hands-on

Sequential Gaussian Simulation Demo
Michael Pyrcz, the University of Texas at Austin, Geostatistical Reservoir Modeling

0. Representative Distribution and Transform	
mean	0.15
st. deviation	0.03

1. Data				
Point	x	y	value	Gaussian
1	70	50	0.1	-1.667
2	5	43	0.14	-0.333
3	18	29	0.2	1.667
unknown	40	30		
mean			0.147	0.00

2. Distance Matrix			
0.00	65.38	56.08	36.06
65.38	0.00	19.10	37.34
56.08	19.10	0.00	22.02

3. Variogram Model			
Nugget	0		
Spherical	1	Range	300

4. Variogram Matrix			
0.000	0.322	0.277	0.179
0.322	0.000	0.095	0.186
0.277	0.095	0.000	0.110

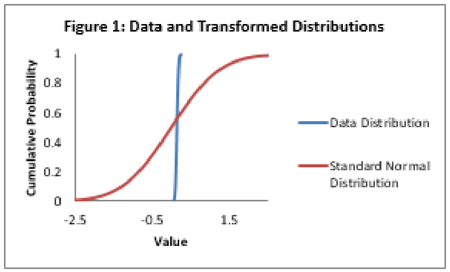


Figure 1: Data and Transformed Distributions

5. Covariance Matrix			
1.000	0.678	0.723	0.821
0.678	1.000	0.905	0.814
0.723	0.905	1.000	0.890

6.Inverse Left Side		
2.109	-0.283	-1.268
-0.283	5.542	-4.809
-1.268	-4.809	6.267

7. Weights	
0.371	
0.000	
0.622	
0.993	*Sum of Weights*

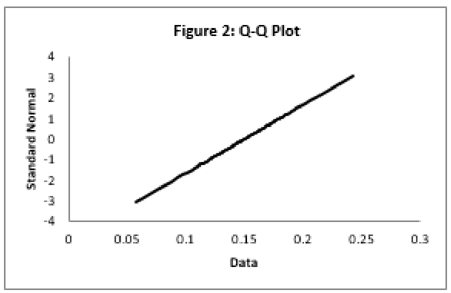


Figure 2: Q-Q Plot

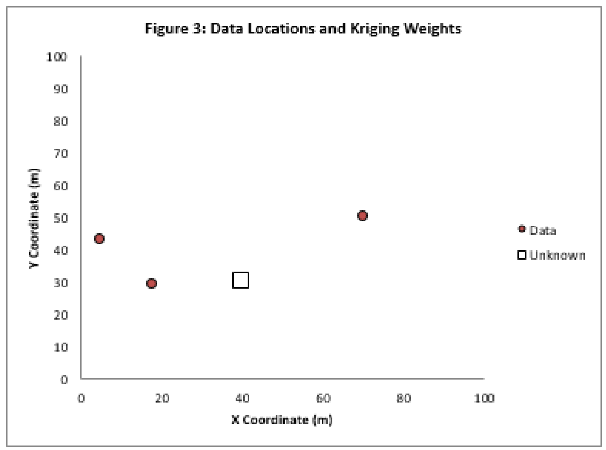


Figure 3: Data Locations and Kriging Weights

8. Kriging Results	
Kriging Estimate	0.417
Kriging Variance	0.142

9. Simulation Results		
Realization	Simulation	Back Transformed
1	0.327	0.160
2	0.806	0.174
3	0.490	0.165
4	0.495	0.165
5	0.043	0.151
6	0.407	0.162
7	-0.109	0.147
8	0.147	0.154
9	0.938	0.178
10	-0.145	0.146

Instructions for Sequential Gaussian Simulation
This sheet provides an illustration of Sequential Gaussian Simulation at a single simulated location.
With the seqential method, the simulated value is added to the data and the next random location is simulated.

Step 0: Model the input distribution. In this case a Gaussian Reference distribution is assumed (see Figure 1). This allows
This allows for the transform of the input data into Gaussian values using the Q-Q plot in Figure 2.
Step 1: Input the data locations and values, the unknown simulated location. At any point these values may be changed to observed their influence on the simulation
Step 2: The distance matrix is calculated, that is the distance between the data and the unknown location.
Step 3: Input the model of spatial continuity in the form of an isotropic spherical variogram and nugget effect (contributions should sum to one). This model may be changed at any time.
Step 4: Variogram matrix is calculated by applying the distance matrix to the isotropic variograms model.
Step 5: Covaraince matrix is calculated by subtracing the variogram from the variance (1 for standard normal distribution).
This is applied to improve numerical stability as a diagonally dominant matrix is more readily invertable.
Step 6: The left hand side of the covariance matrix is inverted.
Step 7: The inverted left handside matrix is multiplied by the right hand side matrix to calculate the simple kriging weights.
Step 8: The kriging estimate and kriging variance are calculated with the weights and covariances.
Step 9: With the Gaussian assumption for the local CDF, the Monte Carlo simulation is applied to draw simulated realizations at the unknown location.

What did we learn?
1. The kriging system under the assumption of Gaussianity provides the full local distribution of uncertainty given local data, previously simulated values, model of spatial continuity and global property distribution.
Note: the multiGaussian assumption is made after univariate transform of the property.
2. Multiple local simulated values are available through Monte Carlo simulation.
3. The variance of the local CDF is indepdent of the loca ldata values (homoscedastic).
4. A variogram sill of 1.0 equal to the standard normal variance is required of the simulation distribution variance will be incorrect.
5. Increased variogram range, decreases the variance of the local CDF.
6. Increased variogram nugget effect, increases the variance of the local CDF and inlcudennt

Gaussian Simulation Hands-on

Some ideas for experimenting with sequential Gaussian simulation. Do the following and pay attention to the weights, the estimate, the estimation variance and simulated values.

1. Set points 1 and 2 closer together.

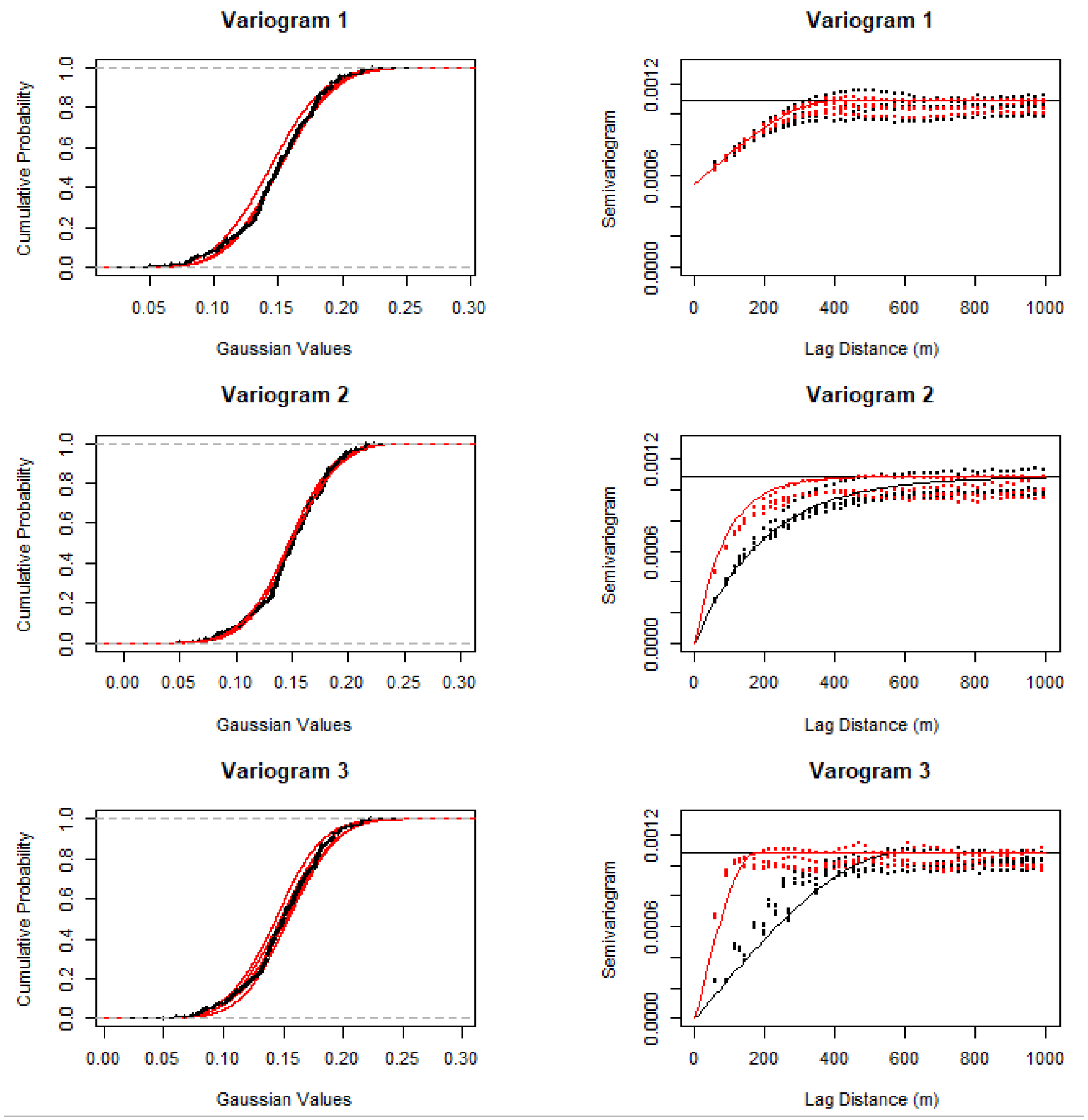
2. Put point 1 behind point 2 to create screening.

3. Put all points outside the range.

4. See the range very large.

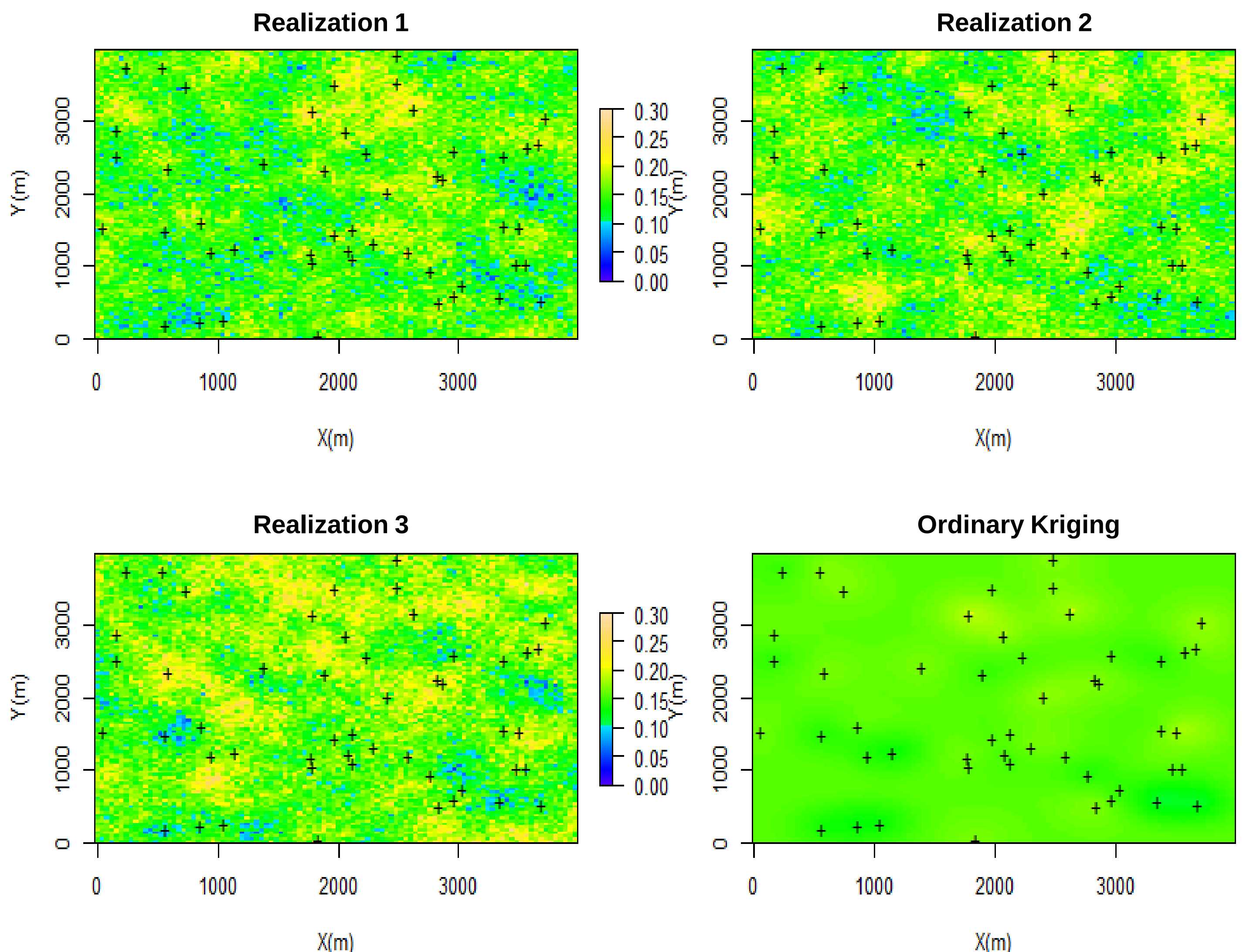
Gaussian Simulation in Excel

- Minimum Acceptance Checks
  - Check for reproduction of the inputs statistics (histogram and variogram).
  - Should be reproduced on average
  - Fluctuations are normal
  - These are proportional to the ratio or variogram range to model size

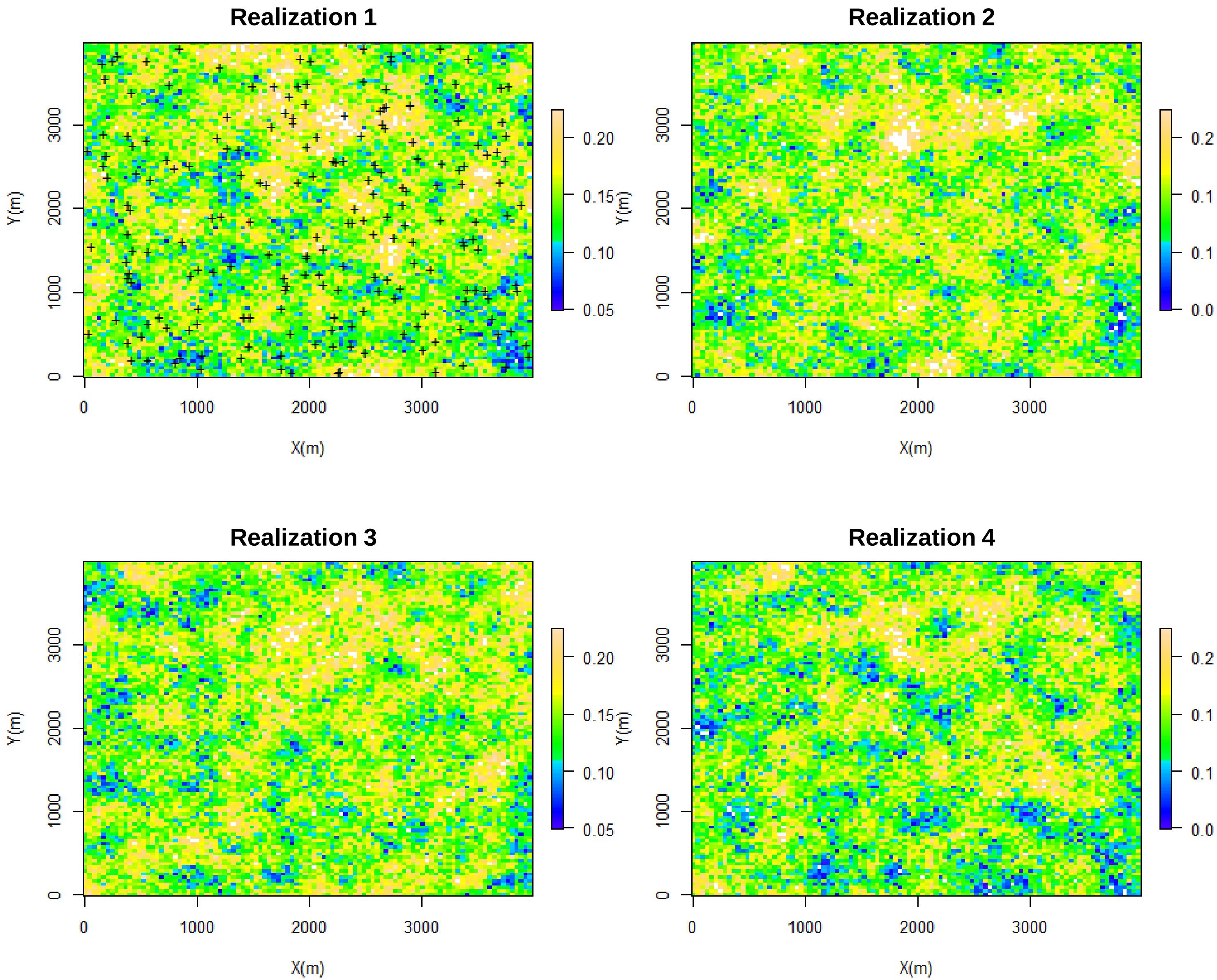


Example of Estimation and Simulation

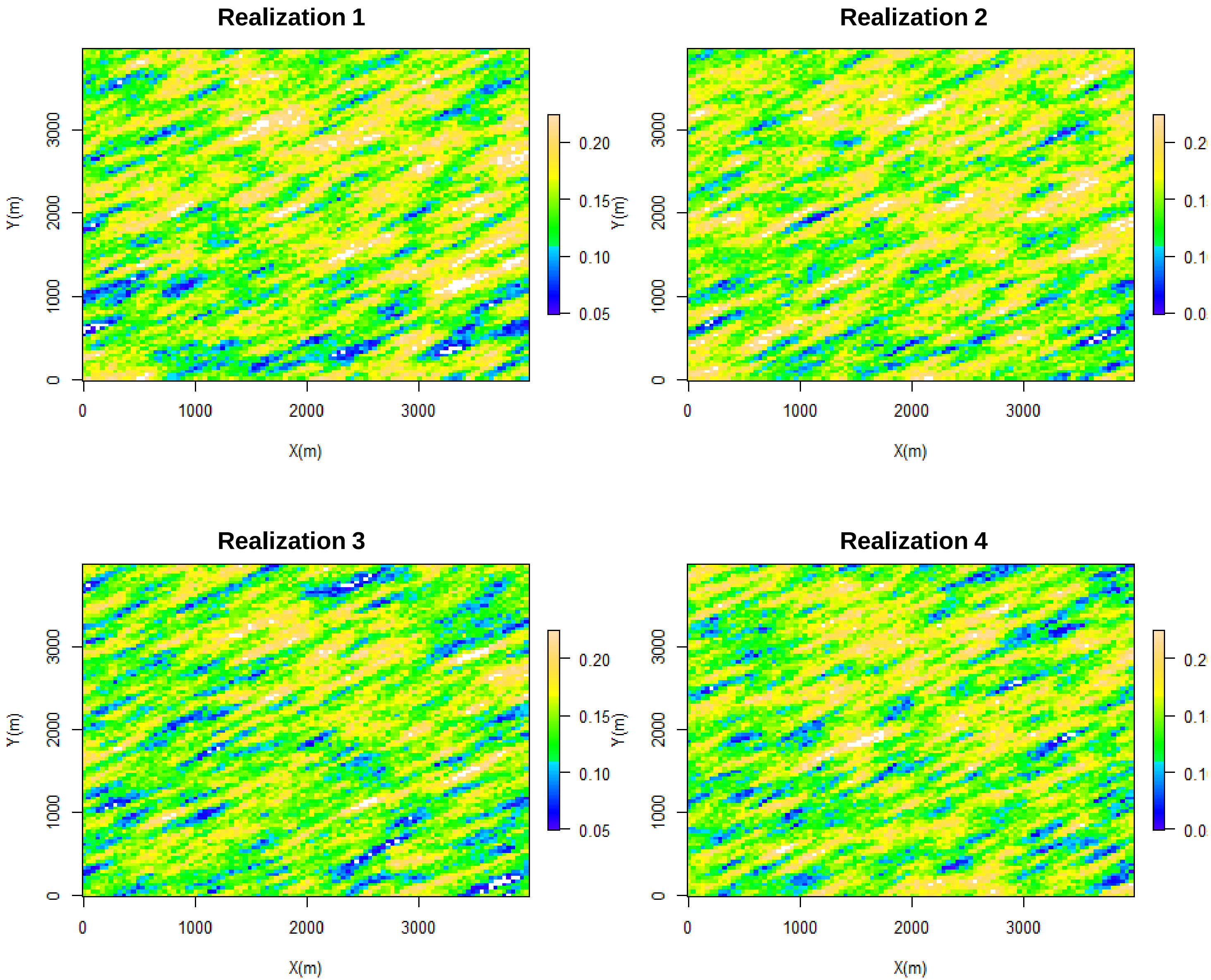
- Compare kriging and simulation.



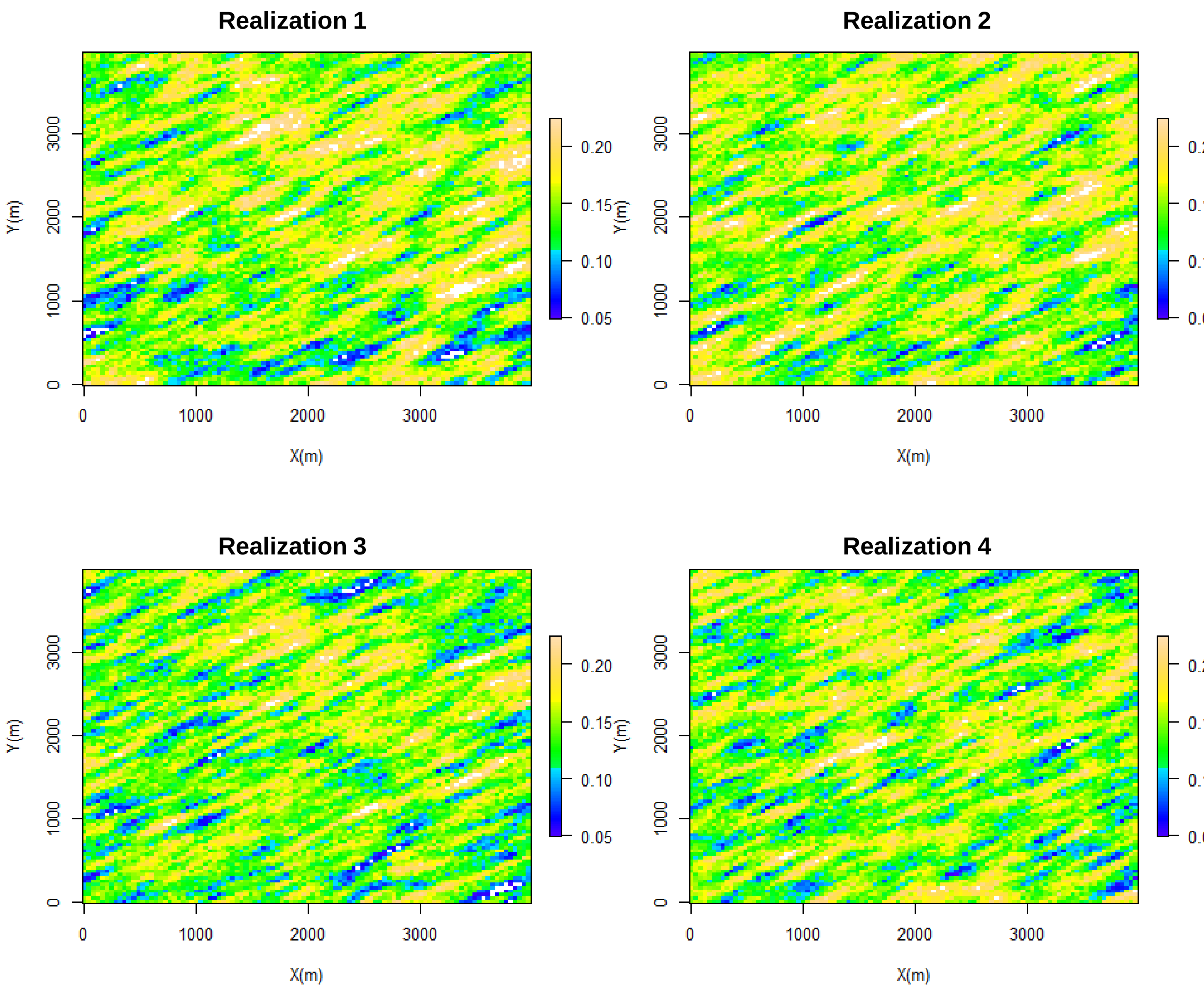
Sequential Gaussian Simulation Demo R with gstat Package



Sequential Gaussian Simulation Demo R with gstat Package



Sequential Gaussian Simulation Demo R with gstat Package



Sequential Gaussian Simulation Hands-on R with gstat Package

- Near the start of the demo, these lines set the 3 variogram models, modify and rerun the realizations and plots.

```
# Anisotropic variogram
vm1 <- vgm(psill = 0.5*sill, "Sph", 400, anis = c(000, 1.0),nugget=0.5*sill)
vm1
vm2 <- vgm(psill = 1.0*sill, "Exp", 200, anis = c(035, 0.5),nugget=0.0*sill)
vm2
vm3 <- vgm(psill = 1.0*sill, "Sph", 600, anis = c(060, 0.2),nugget=0.0*sill)
vm3
```

Some ideas to explore:

1. Set the nugget high (use nugget parameter and set psill to 1-nugget).

2. Set the range even higher lower and higher (currently 200, 400 and 600) in the function call.

3. Increase and decrease the anisotropy ratio (currently 0.5) in the function call.

Simulation
New Tools

Topic	Application to Subsurface Modeling
Predrill with Uncertainty	Estimate at undrilled location *Include uncertainty with a spatial estimate at an unsampled location.*
Simulated Models	Spatial models with some of the spatial features from the actual property of interest *Model realizations with the correct histogram and variogram for subsurface forecasting.*

Spatial Modeling with Geostatistics

Spatial Data Analysis - Modeling

Lecture outline . . .

- **Interpreting Spatial Continuity**
- **Modeling Spatial Continuity**

Prerequisites

Introduction

Probability Theory

Representative Sampling

Spatial Data Analysis

Spatial Estimation

Stochastic Simulation

Uncertainty Management

Machine Learning